IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

ATTORNEY DOCKET NO.: **AT9-99-037**

In re Application of: **CUNG, ET AL.** §

Confirmation No. **8855**

Examiner: **DAY, HERNG DER**

Serial No.: **09/282,619**

Art Unit: **2123**

Filed: **March 31, 1999**

Customer No. **28722**

For: **METHOD FOR COMPUTER MODELS BASED ON ATTRIBUTES SELECTED BY ENTROPY**

APPEAL BRIEF

Commissioner for Patents
Mail Stop Appeal Briefs – Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

This Appeal Brief is submitted in triplicate in support of the Appeal in the above-identified application.

CERTIFICATE OF MAILING
37 CFR 1.8(A)

I hereby certify that this correspondence is being deposited with the United States Postal Service as First Class Mail in an envelope addressed to Commissioner of Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450.

Shenise Ramdeen — 1/16/04

Shenise Ramdeen — Date

REAL PARTY IN INTEREST

The real party in interest in the present application is the Assignee, International Business Machines Corporation of Armonk, New York, as evidenced by the Assignment set forth at Reel 9864/Frame 0552.

RELATED APPEALS AND INTERFERENCES

There are no Appeals or Interferences known to Appellant, the Appellant's legal representative, or assignee, which directly affect or would be directly affected by or have a bearing on the Board's decision in the pending appeal.

STATUS OF CLAIMS

Claims 1, 3-6 and 13-25 stands finally rejected as noted in the Examiner's Action dated August 12, 2003.

STATUS OF AMENDMENTS

No amendment has been submitted subsequent to the final rejection.

SUMMARY OF THE INVENTION

As set forth in the present specification at page 4, lines 15 et seq., attributes of a data set to be employed in generating a predictive model are first analyzed based upon entropy, chi-square, or similar statistical measure. A target group of samples exhibiting one or more desired attributes is identified, then remaining attribute values for the target group are compared to corresponding attribute values for the whole sample population. A subset of all available attributes is then selected from those attributes which exhibit, when comparing attribute values of target group samples to attribute values of the whole sample population, the greatest relative difference or divergence. That is, an attribute for which the target group samples exhibit, for example, only two of all possible values is selected in preference to an attribute for which the target group sample exhibit three or

more possible values. This subset is then employed to generate the predictive model. Efficiency in generating a predictive model is improved utilizing this technique since fewer attributes are employed and less computational resources are required. Accuracy of the resulting predictive model is also improved since attributes potentially skewing the sample population in a manner least related to the desired attribute are eliminated from consideration when developing the model. As illustrated in Figure 3 of the present application and as described at page 11, lines 21 et seq., a high level logic flow chart is depicted for the process of selecting attributes of a sample for generating a predictive model in accordance with a preferred embodiment of the present invention. As illustrated, the process begins at step 302 which depicts a model build being initiated. A data set, from which a sample population may be drawn including at least one sample having a desired attribute, should be available for building the predictive model. If less than the entire data set is employed in generating the predictive model, the resulting predictive model may then be applied to the remaining samples in the data set. The desired attribute (sn) for which the predictive model is generating need not have only two possible values (e.g., gender), but may be a relative measure such as a value exceeding a predetermined threshold.

The process first passes to step 304, which illustrates grouping the elements of the sample population based on the values of the attribute (sn) to be the subject of prediction, identifying a target group of samples. The process then passes to step 306, which depicts selecting an attribute and determining a relative difference or divergence in the attribute values for the target group sample versus the entire sample population. A relative difference (e.g., ratio or percentage) should be determined since comparison of absolute differences may not be meaningful.

The process then passes to step 308, which illustrates a determination of whether all attributes available for the sample population, other than those for which the predictive model is

being built, have been considered. If all attributes for the sample population have not been considered, the process returns to step 306 to select another attribute for analysis and repeats the process of steps 306 with the newly selected attribute.

Once all attributes for the sample population have been analyzed, the process proceeds from step 308 to step 310, which depicts selecting n attributes exhibiting the largest relative differences for samples having the desired attributes as compared to all samples within the sample population. A sort or ranking of the attributes by such relative difference may be useful in this step. The number n of attributes selected may be any arbitrarily set number or, as described above, may be a predetermined percentage of the attributes or attributes exhibiting a relative difference between samples which exceeds a predetermined threshold.

The process next passes to step 312, which illustrates building a model for the desired attribute and the sample population utilizing the selected attributes. Various known techniques may be employed for this purpose. The process then passes to step 314, which depicts applying the predictive model generating to a data set. Finally, the process passes to step 316 which illustrates the process of becoming idol until another model build is undertaken.

The present invention allows data collections having large numbers of potentially irrelevant or meaningless attributes for each sample to be employed in building an accurate predictive model. Efficiency in generating a predictive model is improved by reducing the number of attributes which are considered during the model build. This requires both less time and less computational resources to generate the predictive model. Accuracy of the resulting predictive model also improves since attributes which might skew the sample population but have no relation to the desired characteristic --or less relation to the desired attribute than the other attributes-- are eliminated from consideration in building the predictive model.

ISSUES

1. Is the Examiner's rejection of claims 1, 5-6, 13-15, 18-22 and 25 as unpatentable under 35 U.S.C. § 103(a), over *Piatetsky-Shapiro*, "Discovery, Analysis and Presentation of Strong Rules" (hereinafter referred to as *Piatetsky-Shapiro*), in view of *Simoudis et al.*, United States Patent No. 5,692,107 well founded?

2. Is the Examiner's rejection of claims 3-4, 16-17 and 23-24 as unpatentable under 35 U.S.C. § 103(a), over the combined teachings of *Piatetsky-Shapiro*, *Simoudis et al.* and further in view of *Dash et al.*, "Dimensionality Reduction of Unsupervised Data" (hereinafter referred to as *Dash et al.*), well founded?

GROUPING OF THE CLAIMS

For purposes of this Appeal, claims 1, 5-6, 13-15, 18-22 and 25 stand or fall together as a first group, claims 3-4, 16-17 and 23-24 stand or fall together as a second group.

ARGUMENT

The Examiner has rejected claims 1, 5-6, 13-15, 18-22 and 25 under 35 U.S.C. § 103(a) as being unpatentable over *Piatetsky-Shapiro*, in view of *Simoudis et al.* That rejection is not well founded and should be reversed.

The Examiner relies upon *Piatetsky-Shapiro* to teach the claim limitation of "comparing said one or more desired attributes and respective values with said sample population to obtain a target population." First, Appellants point out that the reference is devoid of any teaching for obtaining a target population. The "target population" is an important claim element as a statistical measure of difference between attributes and respective values in the "target population" as compared to the sample population to "reducing the number of attributes and respective values of the sample population."

The Examiner believes that "*Piatetsky-Shapiro* expressly teaches obtaining a target population in the last two lines of page 235" which recite:

At the end, a cell for A = a contains the summary of all the file tuples satisfying A = a. The summary can be presented to the user or used for deriving rules implied by A = a.

Appellants contend the above lines only indicate that the result of the KID3 algorithm taught by *Piatetsky-Shapiro* produces a summary of the existing sample population, and not obtaining a target population. Support for Appellants interpretation may be found at page 235 of *Piatetsky-Shapiro* which recites: "I present here the KID3 algorithm that finds, in parallel, all simple exact rules of the form (A = a) --> cond(Bi)" and "... the cell summary is updated ..." Appellants contend that a summary of an existing sample population is not the obtaining of a target population.

Second, *Piatetsky-Shapiro* does not teach or suggest determining a statistical measure of difference between each of the attributes and respective values of the target population and sample population as recited in Claim 1. In the claimed invention, the selected target population is compared to the entire sample population to determine which attributes and respective values are most likely relevant in computing a predictive model. The comparison of the obtained target population to the sample population yields different results than simply reducing a data set to a set of rules as in *Shapiro*. The results depend on the selected target group and not the population as a whole. Different target groups may result in a different selection of most relevant attributes. For example, a target group for the purchase of a type of pizza may show a strong correlation with age and no other attribute while the target group for the purchase of an expensive product may show a correlation with income.

The Examiner asserts that *Simoudis et al.* teaches the selection of a data analysis module to perform data mining, including the use of a target population Appellants acknowledge that *Simoudis et al.* teaches the use of a target population that is employed in generating a predictive model. However, *Simoudis et al.* does not teach "comparing said one or more desired attributes and respective values with said sample population to obtain a target population" as recited by the claims in the present invention. *Simoudis et al.* only teaches that the target data set typically represents a subset of a larger underlying data source and may be compiled from sources with difference data formats (Col. 4 lines 16-17). The present invention teaches a technique, not found in the prior art, for selecting a target group by comparing attributes values of the sample population to desired values and reducing the number of attributes by determining the statistical measure of difference between the attributes of the target and sample populations.

The Examiner has also rejected claims 3-4, 16-17 and 23-24 under 35 U.S.C. § 103(a) as being unpatentable over the combined teachings of *Piatetsky-Shapiro*, *Simoudis et al.* and *Dash et al.*, "Dimensionality Reduction of Unsupervised Data", Proceedings, Ninth IEEE International Conference on Tolls with Artificial Intelligence, Nov. 1997, this rejection is not well founded and it should be reversed.

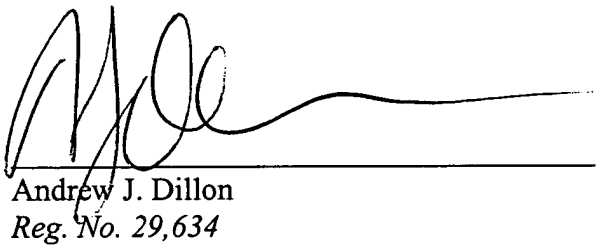
Dash et al., teaches a dimensionality reduction of unsupervised data utilizing an entropy measure. A sequential backwards selection algorithm, SUD, is implemented within *Dash et al.* to determine the relative importance among features by determining the relevance of particular features. *Dash et al.* are entirely silent on the subject of the reduction of variables based upon a difference between the attributes and the respective values of a target group and sample population. Consequently, Applicant urges that *Dash et al.*, whether considered alone or in combination with *Piatetsky-Shapiro* and *Simoudis et al.* fails to teach or suggest in any way each of the claim limitations of the present application. Specifically, these combined citations lack any teachings or suggestions of a determination of a statistical measure of difference between the attributes and the respective values of a target population and a sample population or the comparing of attributes and respective values with a sample population to obtain a target population. Consequently, Applicant urges that the rejection of the claims of group two is not well founded and it should be reversed.

For the rejections under 35 U.S.C. § 103(a) to be well founded, the Examiner must present prior art which teaches or suggests every limitation of the claimed (sn) rejection. The combination of *Piatetsky-Shapiro, Simoudis et al.* and *Dash et al.*, whether considered singly or together do not teach or suggest every claim limitation of the present invention. Most notably, the cited prior art lacks any teaching of determining of a statistical measure of difference

between attributes and the respective values of an obtained target population and a sample population or comparing attributes and respective values with a sample population to obtain a target population. Accordingly, Applicant urges that all rejections in this application are not well founded and should be reversed.

Please charge the fee of $330.00 for submission of a Brief in Support of Appeal to IBM Corporation Deposit Account No. 09-0447. No additional filing fee is believed to be necessary; however, in the event that any additional fee is required, please charge it to IBM Deposit Account Number 09-0447.

Respectfully submitted,



Andrew J. Dillon
Reg. No. 29,634
BRACEWELL & PATTERSON, L.L.P.
P.O. Box 969
Austin, Texas 78767-0969
512.542.2121

ATTORNEY FOR APPLICANT

APPENDIX

1. A method of reducing the number of the number of attributes and respective values of a sample population employed in generating a predictive model, said method comprising the steps of:

obtaining one or more desired attributes and respective values;

comparing said one or more desired attributes and respective values with said sample population to obtain a target population;

determining a statistical measure of difference between each of the attributes and respective values of said target population and the attributes and respective values of the sample population; and

utilizing said statistical measure of difference to reduce the number of attributes and respective values of said sample population.

2. (Cancelled)

3. The method of claim 1, wherein the step of determining a statistical measure of difference further comprises:

determining an entropy for the attribute values.

4. The method of claim 1, wherein the step of utilizing said statistical measure to reduce the number of attributes and respective values of said population further comprises:

identifying n attributes having a largest difference in respective values with said target population.

5. The method of claim 1, wherein the step of utilizing said statistical measure to reduce the number of attributes and respective values of said population further comprises:

identifying a predetermined percentage of attributes and respective values having a larger statistical measure of difference than remaining attributes and respective values.

6. The method of Claim 1, wherein the step of utilizing said statistical measure to reduce the number of attributes and respective values of said population further comprises:

identifying attributes and respective values where said statistical measure of difference exceeds a predetermined amount.

7-12. (Cancelled)

13. A method of selecting attributes for computing a model, comprising:

for a plurality of samples each having values for a plurality of attributes:

for each of the plurality of attributes:

comparing the attribute values for a target group of samples to the attribute values for all of the plurality of samples; and

determining a difference between the attribute values for the target groups and the attribute values for all of the plurality of samples; and

identifying attributes within the plurality of attributes having a largest difference between the attribute values for the target groups and the attribute values for all of the plurality of samples; and

selecting at least some of the identified attributes.

14. A system for selecting attributes for computing a model, comprising:

a memory containing data for a plurality of samples each having values for a plurality of attributes; and

a processor coupled to the memory and executing a selection process including:

comparing attribute values for samples having a desired attribute value to attribute values for all samples;

selecting a subset of available attributes based on a difference between attribute values for the samples having the desired attribute value and attribute values for all of the samples; and

employing the selected subset of attributes to generate a predictive model.

15. The system of claim 14, wherein the selection process determines a statistical measure of difference between the attribute values for samples having the desired attribute and the attribute values for all of the samples.

16. The system of claim 15, wherein the selection process determines an entropy for the attribute values.

17. The system of claim 14, wherein the selection process identifies a predetermined number of attributes having a largest difference in the attribute values for selection.

18. The system of claim 14, wherein the selection process identifies a predetermined percentage of attributes having a larger difference in the attribute values for selection.

19. The system of claim 14, wherein the selection process identifies, for selection, attributes having a difference in the attribute values exceeding a predetermined amount.

20. A system for computing a model, comprising:

a memory containing data for a plurality of samples each having values for a plurality of attributes; and

a processor coupled to the memory and executing a selection process including:

comparing attribute values for a target subset of the plurality of samples to attribute values for all of the samples;

selecting attributes having a largest difference between attribute values for the target subset and attribute values for all of the samples; and

computing a model employing the selected attributes.

21. A computer usable medium for selecting attributes for computing a model, said computer usable medium comprising:

computer program code for reading values of attributes for a plurality of samples;

computer program code for comparing attribute values for samples having a desired attribute value to attribute values for all samples; and

computer program code for selecting a subset of available attributes based on a difference between attribute values for samples having the desired attribute value and attribute values for all samples.

22. The computer usable medium of claim 21, wherein the computer program code for comparing attribute values for samples having a desired attribute value to attribute values for all samples further comprise:

computer program code for determining a statistical measure of difference between the attribute values for samples having the desired attribute value and the attribute values for all samples.

23. The computer usable medium of claim 22, wherein the computer program code for determining a statistical measure of difference between the attribute values for samples having the desired attribute value and the attribute values for all samples further comprise:

computer program code for determining an entropy of the attribute values for samples having the desired attribute value and an entropy of the attribute values for all samples;

computer program code for comparing the entropy of the attribute values for samples having the desired attribute value to the entropy of the attribute values for all samples for each attribute to determine a relative measure of difference; and

computer program code for comparing the relative measure of difference of all attributes.

24. The computer usable medium of claim 21, wherein the computer program code for selecting a subset of available attributes based on a difference between attribute values for samples having the desired attribute value and attribute values for all samples further comprise:

computer program code for identifying n attributes having a largest difference in the attribute values.

25. A computer usable medium for selecting attributes for computing a model, said computer usable medium comprising:

computer program code for comparing attribute values for a target group of samples to attribute values for all samples for each of a plurality of attributes;

computer program code for determining a difference between the attribute values for the target group of samples and the attribute values for all of the samples; and

computer program code for selecting a group of attributes having a largest difference between the attribute values for the target group of samples and the attribute values for all samples.